Environmental variables affect the observed to expected heterozygosity ratio in human populations

Fumagalli M[1,2], Pozzoli U[1], Pattini L[2], Bresolin N[1,3], Sironi M[1].

[1] Scientific Institute IRCCS E. Medea, Bioinformatic Lab, 23842 Bosisio Parini (LC), Italy;
[2] Bioengineering Department, Politecnico di Milano, 20133 Milan, Italy;
[3] Dino Ferrari Centre, Department of Neurological Sciences,
University of Milan, IRCCS Ospedale Maggiore Policlinico,
Mangiagalli and Regina Elena Foundation, 20100 Milan, Italy.

Motivation

Patterns of human genetic diversity in modern populations are the results of different factors including demographic (e.g. migrations, population size changes) and evolutionary events (i.e. natural selection). Discovering signatures of natural selection could lead to highlight how different populations have adapted to their environment over evolutionary history. Although neutral migratory events, starting from the human out-of-Africa spread, explain a main fraction of worldwide cline in nucleotide diversity, genetic adaptation to local variation in pathogens, climate and diet has been an important step during human evolution. Given that infectious diseases have represented one of the major threats to human populations, it's conceivable that parasites have been acting as a powerful selective force. Pathogen-driven selection operates when specific alleles are favored due to their ability to improve resistance to different parasite species. Examples of overdominant selection, a case of balancing selection where heterozygotes have higher fitness than homozygotes, imposed by pathogens have been previously reported for loci involved in immune response. Given these premise, we wished to investigate the relationship between genetic diversity and the richness of different pathogen species in human populations distributed worldwide.

Methods

We exploited the availability of CEPH Human Genome Diversity Panel, comprising more than 1000 individuals from widely distributed populations typed with a large number of markers. Genotype data for this panel were retrieved from a previous work. Parental individuals and populations with low sample size, as well as markers falling within Copy Number Variation regions were removed, leaving nearly 495k SNPs typed in 947 individuals grouped in 37 populations. Pathogen absence/presence matrices for the 21 countries where HGDP-CEPH populations are located were derived from the Gideon database. Climatic variables were derived from the NCEP/NCAR database. For each population we calculated two measures of nucleotide diversity: the expected heterozygosity (an index of neutral genetic diversity) and the ratio between the observed and the expected heterozygosity (an index revealing an excess or loss of heterozygotes relative to neutral expectations). In order to assess the relationship between heterozygosity and environmental variables, we fitted a multiple linear regression model which incorporates the distance from Central East Africa (as a descriptor of migratory events). The significance of the different variables was assessed using partial F-test. We ran Principal Component Analysis on 8 climatic variables representing information about temperature, precipitation rate and short wave irradiation for each geographic location.

Results

Variation in genetic diversity in any population is a function of both demography and selective forces. Since it has been shown that geographic distance through landmasses from Central East Africa is well correlated with the expected heterozygosity, this measure can be used to disentangle the effect of past colonization routes. Regression analyzes for the relationship among nucleotide diversity, distance from Africa and pathogen species richness reveal that both neutral and selective events (as a results of adaptation to local environment) have shaped human genetic variation, as previously suggested. On the other hand, pathogen richness is the only significant predictor for the ratio of observed to expected heterozygosity; in particular, a positive correlation between pathogen richness and high heterozygosity was observed. These data indicate that pathogens have exerted a significant selective pressure on human genetic diversity. With the aim of taking into account the contribution of climatic factors to the observed correlation, 8 different climatic variables were reduced in dimension by Principal Component Analysis with the first 2 components explaining 84% of total variance. While both components are well correlated with neutral genetic diversity, only the second one (explaining nearly 30% of total variance mainly composed by maximum and medium precipitation rate and maximum annual temperature) shows a significant linear relationship with the ratio of observed to expected heterozygosity. Moreover, this climatic component is highly correlated with pathogen richness. Given previous reports on specific immune response loci and the correlation of climatic variables with pathogen richness, this latter might be regarded as the principal factor explaining the worldwide variation in the observed/expected heterozygosity ratio. These findings, other than restate the importance of human adaptation to local environment, allow the identification of genetic variants shaped by natural selection which can be thought as possible candidates for holding a functional role, specially regarding to modulation of susceptibility or protection to common infectious disease.

Contact : matteo6.fumagalli@mail.polimi.it